



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/559,571	12/05/2005	Yasushi Sato	0670-7064	8897
31780	7590	10/09/2009	EXAMINER	
ERIC ROBINSON			LERNER, MARTIN	
PMB 955			ART UNIT	
21010 SOUTHBANK ST.			PAPER NUMBER	
POTOMAC FALLS, VA 20165			2626	
			MAIL DATE	DELIVERY MODE
			10/09/2009	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary

Application No.

10/559,571

Applicant(s)

SATO, YASUSHI

Examiner

MARTIN LERNER

Art Unit

2626

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --
Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 10 September 2009.
2a) ☒ This action is **FINAL**. 2b) ☐ This action is non-final.
3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 23 to 40 is/are pending in the application.
4a) Of the above claim(s) _____ is/are withdrawn from consideration.
5) ☐ Claim(s) _____ is/are allowed.
6) ☒ Claim(s) 23 to 40 is/are rejected.
7) ☐ Claim(s) _____ is/are objected to.
8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
10) ☐ The drawing(s) filed on _____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☒ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
a) ☒ All b) ☐ Some * c) ☐ None of:
1. ☒ Certified copies of the priority documents have been received.
2. ☐ Certified copies of the priority documents have been received in Application No. _____.
3. ☒ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- 1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-946)
3) ☐ Information Disclosure Statement(s) (PTO/CDC)
4) ☐ Interview Summary (PTO-413)
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____
Paper No(s)/Mail Date _____

DETAILED ACTION

Claim Rejections - 35 USC § 103

1. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

2. Claims 23, 35, and 37 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Reynar et al.* in view of *Itoh*.

Concerning independent claims 23, 35, and 37, *Reynar et al.* discloses a speech synthesis device, method, and computer program, comprising:

“a first storage means for storing a plurality of pieces of voice unit data representative of one or more speech words” – stored audio data 270 is a long-term storage medium for converting speech input 290 from a speech recognition program 240; stored audio data 270 may later be accessed for audio playback (column 9, lines 5 to 10: Figure 2);

“a selection means for selecting voice unit data whose reading is common with a speech word composing inputted sentence information from the plurality of pieces of voice unit data stored in the first storage means” – if multi-source input and playback utility 200 determines that stored audio data 270 is linked to a word, then the utility retrieves this audio data; a user selects a text portion of a document which he desires the multi-source input and playback utility to play; the multi-source input and playback

utility 200 determines whether the word is linked to stored audio data 270 saved from a previous dictation session (column 11, lines 33 to 56: Figure 4: Steps 410 and 415);

“a missing part synthesis means, for a speech word among the sentence information for which the selection means could not select the voice unit data, for synthesizing speech data representative of a desired speech waveform” – alternately, utility 200 may determine that no speech is linked to the word; in this event, the utility checks for the existence of a TTS entry 220 corresponding to the current word; if such a TTS entry 220 exists, the TTS module 137 retrieves the TTS entry and returns it to the word processor 210 (column 12, lines 9 to 27: Figure 4: Steps 410, 425, 430, and 440);

“a synthesis means for combining the voice unit data selected from the selection means and the speech data synthesized by the missing part synthesis means to create data representative of a synthesis speech corresponding to the sentence information” – word processor 210 parses each word within the text selection in turn, and retrieves and plays either stored audio data 270 or a TTS entry 220; to a user of the multi-source input and playback utility 200, a continuous stream of mixed stored audio data and TTS entries is heard, sounding out the text selection (column 10, lines 43 to 50: Figure 2);

“wherein the missing part synthesis means has a second storage means for storing a plurality of pieces of data representative of one or more pitches of voice waveform fragments” – optionally, the audible characteristics of the TTS entry 220, such as pitch, tone, and speech, may be manipulated by the utility prior to playback in order to more closely match the sound of the TTS entry to that of the stored audio data (column 12, lines 31 to 35: Figure 4); implicitly, an TTS entry will have at least “one

pitch", which can then be manipulated prior to playback, and is stored in a TTS entry database 220 ("a second storage means") (Figure 2);

"wherein data representative of voice waveform fragments composing the speech word whose voice unit data could not be selected is acquired from the second storage means and the acquired data is mutually combined to synthesize the speech data representative of the desired speech waveform" – word processor 210 parses each word within the text selection in turn, and retrieves and plays either stored audio data 270 or a TTS entry 220; to a user of the multi-source input and playback utility 200, a continuous stream of mixed stored audio data and TTS entries is heard, sounding out the text selection (column 10, lines 43 to 50: Figure 2).

Concerning independent claims 23, 35, and 37, the only element not expressly disclosed by *Reynar et al.* is "the one or more pitches of voice waveform fragments being cut off in a unit of voice pitch from an actual speech waveform". *Reynar et al.* discloses that TTS (text-to-speech) entries can be manipulated by pitch, but does not say that the TTS entries are synthesized in units of pitch. Still, it is fairly well known in text-to-speech synthesis that it is advantageous to synthesize speech in units of pitch for voiced segments to make it easier to splice together the waveforms. Specifically, *Itoh* teaches text-to-speech synthesis by concatenation of waveform segments, where a representative phoneme waveform is cut out or sliced every fixed period. When a phoneme waveform is a voiced sound, the waveform is cut out every fundamental period, which cutout is called a pitch synchronous cutout. When a waveform is cut out for every fundamental period, *i.e.*, every pitch period for voiced sounds, pitch marking

part 17 detects a pitch period and a mark indicating the reference position of the speech pitch period -- what is called a pitch mark -- is added to the waveform information.

(Column 8, Line 52 to Column 9, Line 13: Figure 6) An objective is to make improvements to conventional waveform compilation type speech synthesis which permits synthesis of more natural and smooth speech. (Column 3, Lines 4 to 9) It would have been obvious to one having ordinary skill in the art to cut out voiced waveform fragments in units of pitch as taught by *Itoh* for the TTS entries of *Reynar et al.* for a purpose of making improvements to conventional waveform concatenation so as to synthesize speech in a manner that is more natural and smooth.

3. Claims 24 to 29, 34, 36, and 38 to 40 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Reynar et al.* in view of *Itoh* as applied to claim 23 above, and further in view of *Kato et al.* (EP '072).

Concerning independent claims 34, 36, and 38, *Reynar et al.* discloses a speech synthesis device, method, and computer program, comprising a first storage means, a selection means, and a missing part synthesis means of independent claims 23, 35, and 37, but does not expressly disclose the limitations of "wherein the first storage means stores phonetic data representative of a reading of the voice unit data with the phonetic data being associated with the voice unit data, and wherein the selection means operates to handle voice unit data which is associated with the phonetic data representative of a reading matching with the reading of a speech word composing the sentence information as voice unit data whose reading is common with the speech

word." *Reynar et al.* suggests that stored audio data 270 corresponds to phonetic data from speech recognition. (Column 9, Lines 2 to 10) However, *Reynar et al.* doesn't disclose choosing a voice unit from a plurality of alternative voice units that is associated with a desired reading of a speech word in the context of a sentence. Still, *Kato et al.* (EP '072) teaches a speech synthesizing system and speech synthesizing method, where speech synthesis is performed taking into account a construction of a sentence. (§§0007 - §§0008) A prosodic data retrieving section 140 searches prosodic data stored in prosodic information database 130 in response to output from language processing section 120, and outputs the search result. The retrieval keys that match the search key to a certain degree are selected as retrieval candidates, and of the selected candidates, the key having the highest degree of matching is selected. (§§0062 - §§0063) Prosodic information corresponds to "phonetic data representative of a reading of the voice data unit". An objective is to provide a speech synthesis system capable of generating natural sounding speech from arbitrary input texts having good sound quality. (§§0009) It would have been obvious to one having ordinary skill in the art to store phonetic data representative of a reading of the voice unit data so as to match a reading of a word in a sentence by prosody as taught by *Kato et al.* (EP '072) in a multi-source input and playback utility of *Reynar et al.* for a purpose of generating natural sounding speech having good sound quality.

Concerning claims 24 to 27 and 39, *Kato et al.* (EP '072) teaches matching prosody. (§§0062 - §§0063) Prosody is equivalent to "cadence". The retrieval keys that match the search key to a certain degree are selected as retrieval candidates, and

of the selected candidates, the key having the highest degree of matching is selected. Implicitly, those candidates that do not have the highest degree of matching are excluded ("to exclude from the objects of selection voice unit data whose cadence does not match with the cadence prediction result under the predetermined conditions").

Concerning claims 28 to 29 and 40 to 41, *Reynar et al.* discloses audio characteristics include pitch, tone, and speed. (Column 12, lines 31 to 35) *Kato et al.* (EP '072) teaches that prosodic information database 130 stores a fundamental frequency pattern, and prosodic data retrieval section 140 retrieves a fundamental frequency pattern having the highest match. (¶¶0062] - ¶¶0063]) Prosody is equivalent to "cadence", and a fundamental frequency pattern corresponds to "a time variation in pitch" because the fundamental frequency is the same as "pitch", and the pattern corresponds to its time evolution. See Figures 2 to 4 of *Kato et al.* (EP '072).

4. Claims 30 to 33 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Reynar et al.* in view of *Itoh* and *Kato et al.* (EP '072) as applied to claims 23 to 25 above, and further in view of *Chihara*.

Reynar et al. discloses audio characteristics include pitch, tone, and speed. (Column 12, lines 31 to 35) However, *Reynar et al.* suggests manipulating the speed characteristics of a TTS entry, but does not expressly say that utterance speed conversion means acquires utterance speed data specifying conditions, selecting or converting speech data and/or voice unit data at a speed fulfilling the specified conditions, and eliminating or adding segments by the utterance speed conversion

means. Still, *Chihara* teaches a method of controlling high-speed reading in a text-to-speech conversion system, where control factors are required to predict a duration length of each phoneme or word. The prediction uses pieces of information such as the phoneme, the kind of adjacent phonemes, the number of mora in the phrase, and the position in the sentence, which are sent to a duration estimation section. The predicted result is sent to a duration correcting section to correct the predicted value where the user designates the utterance speed. (Column 5, Lines 34 to 67: Figure 20) At a high utterance speed, a number of superimposed voice segments is subtracted ("by eliminating a segment") to make the waveform, and at a low utterance speed, the number of superimposed segments is repeated ("adding a segment") for making the waveform. (Column 6, Lines 1 to 11: Figure 21) An objective is to control reading speed from a phoneme and prosody character string including accent and intonation. (Column 1, Lines 19 to 28: Figure 15) It would have been obvious to one having ordinary skill in the art to provide utterance speed conversion at a speed fulfilling specified conditions as taught by *Chihara* in a multi-source input and playback utility of *Reynar et al.* for a purpose of controlling reading speed from a prosody character string including accent and intonation.

Response to Arguments

5. Applicant's arguments filed 10 September 2009 have been considered but are moot in view of the new grounds of rejection, necessitated by amendment.

Applicant has noted several errors in the rejection, which are being corrected herein. Specifically, Applicant points out an error on Form PTOL-326 that claims 1 to 41 are pending and rejected, whereas, in fact, claims 1 to 22 were cancelled by Preliminary Amendment filed on 22 February 2006. The Examiner agrees that this is the case, and that Form PTOL-326 should have stated that claims 23 to 41 were pending and rejected. However, the error only appears to be one of a simple typographical nature because the substantive rejection following on Pages 2 to 10 of the Office Action, in fact, only rejects claims 23 to 41.

Moreover, Applicant has pointed out that the Office Action fails to include a citation and copy of *Kato et al. (EP '072)*. The Examiner apologizes for this oversight. Applicant notes that the Office Action cites, as supplemental prior art, a corresponding equivalent patent as issued in the United States, *Kato et al. ('309)*. *Kato et al. (EP '072)* is relied upon in a rejection of the pending claims. Applicant has respectfully requested that *Kato et al. (EP '072)* be properly cited on Form PTO-892, and that a copy be included in a subsequent communication, if the rejection continues to rely on *Kato et al. (EP '072)*.

The Examiner apologizes for this oversight in failing to provide a copy of *Kato et al. (EP '072)*. A proper citation and copy are now included for *Kato et al. (EP '072)*. However, it appears that Applicant was not substantially prejudiced by the omission due to the citation of the corresponding equivalent patent, *Kato et al. ('309)*, filed in the United States. Nor has Applicant made a timely request for a supplemental Office Action to correct the failure to cite *Kato et al. (EP '072)*. Thus, it is believed that the

finality of the current Office Action is proper, as necessitated by amendment, as Applicant does not appear to be prejudiced by the omission.

Applicant's arguments directed to the amendments of independent claims 23, 35, and 37 are moot. Admittedly, *Reynar et al.* does not teach the claim limitation of "the one or more pitches of voice waveform fragments being cut off in a unit of voice pitch from an actual speech waveform". However, that limitation is taught by *Itoh* in a fairly well known technique for making speech sound more natural when waveform segments are concatenated together.

Applicant presents one argument deserving of comment. Applicant states that a voice waveform fragment generally has an extremely short duration time of approximately 1 to 3 ms as compared to a duration time for any phoneme of approximately 100 to 400 ms, so that Applicant's technique combines voice waveform fragments at voice pitch units having an extremely short duration time, which is different from than *Reynar et al.*, which uses voice waveforms having a relatively large duration time.

However, Applicant's Specification does not appear to disclose any numerical estimates for voice fragment durations, nor is any claim directed to a numerical pitch duration. Typically, a pitch of the human voice is in the range of 50 Hz to 1000 Hz, and a subframe/frame of speech has a length of 5 ms/30 ms. Generally, a pitch period may be anywhere between 1 ms to 15 ms because a low fundamental frequency of about 100 Hz should correspond to a pitch period of about 10 ms. Applicant's analysis may somewhat underestimate the pitch period of a human voice. Still, independent claims

23, 35, and 37 say that the pieces of data representative of the pitch of voice fragments that are cut off is "one or more pitches". Similarly, *Itoh* shows how a stored waveform segment for concatenation is cut out and marked over a number of pitch periods. Thus, the claim limitations appear to be met even if more than one pitch period is representative of a waveform, as long as at least one marked pitch period is present in a stored waveform segment.

Conclusion

6. The prior art made of record and not relied upon is considered pertinent to Applicant's disclosure.

Kamai et al. ('812) discloses related prior art directed to synthesizing speech by pitch waveforms.

Huang et al., Chu et al., Vermeulen et al., Nukaga et al., Holm et al., and Kato et al. ('451) disclose related prior art.

7. Applicant's amendment necessitated the new grounds of rejection presented in this Office Action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the

shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to MARTIN LERNER whose telephone number is (571)272-7608. The examiner can normally be reached on 8:30 AM to 6:00 PM Monday to Thursday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, David R. Hudspeth can be reached on (571) 272-7843. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a

Art Unit: 2626

USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Martin Lerner/
Primary Examiner
Art Unit 2626
October 8, 2009